

Online High-Accurate Calibration of RGB+3D-LiDAR for Autonomous Driving

Tao Li¹, Jianwu Fang^{1,2}, Yang Zhong¹, Di Wang¹, and Jianru Xue¹(✉)

¹ Laboratory of Visual Cognitive Computing and Intelligent Vehicle,
Xi'an Jiaotong University, Xi'an, People's Republic of China
jrxue@mail.xjtu.edu.cn

² School of Electronic and Control Engineering,
Chang'an University, Xi'an, People's Republic of China

Abstract. Vision+X has become the promising tendency for scene understanding in autonomous driving, where X may be the other non-vision sensors. However, it is difficult to utilize all the superiority of different sensors, mainly because of the heterogenous, asynchronous properties. To this end, this paper calibrates the commonly used RGB+3D-LiDAR data by synchronization and an online spatial structure alignment, and obtains a high-accurate calibration performance. The main highlights are that (1) we rectify the 3D points with the aid of differential inertial measurement unit (IMU), and increase the frequency of 3D laser data as the same as the ones of RGB data, and (2) this work can online high-accurately updates the external parameters of calibration by a more reliable spatial-structure matching of RGB and 3D-LiDAR data. By experimentally in-depth analysis, the superiority of the proposed method is validated.

1 Introduction

Calibration of multiple sensors is the prerequisite of robust fusing them. The coexistence of multiple sensor has become the standard configuration for mobile robot systems, especially for autonomous driving. Based on the investigation for sensor utilization for autonomous driving [1], in addition to the visual camera system when perception [13, 14], other non-vision sensors play very important role for scene understanding, even with a more emphasis. The main reason is that vision camera is vulnerable to dynamic environment, but with a stable perception by other non-vision ones, such as laser range finders. However, vision camera has the most intuitive and richest representation for scene knowledge. Therefore, calibration for multiple sensors is an inevitable work for effectively scene perception. However, the calibration of camera and a laser range finder is challenging because of the different physical meaning of data gathered by distinct sensors [4]. In addition, the frame frequency and point sparsity of two sensor planes are of disparity significantly, seeing Fig. 1 for a demonstration.

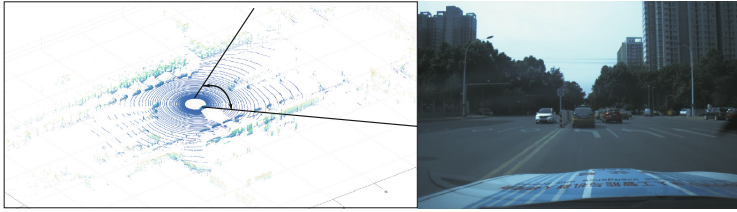


Fig. 1. One frame of RGB and 3D-LiDAR data.

1.1 Related Works and Limitations

Facing this issues, many works carried out the calibration for RGB and Laser data owing to that these two kind of sensors are the most common in the existing mobile robot systems [5, 8]. To summarize the calibration methods, they can be divided into two categories: hand-operated calibration and automatic calibration.

Hand-operated calibration: Hand-operated calibration for a long time is the main module of calibration for different sensors. That is because the calibration with an human auxiliary has a controllable condition for better key point detection and correlation, by which better extrinsic calibration parameters can be obtained. For the early calibration works, the chessboard is usually utilized for a better keypoint detection. For example, Zhang and Pless [11] proposed a calibration method by adding the planar constraints of camera and Lasers with a manual selection of chessboard region. Unnikrishnan and Hebert [10] designed a toolbox for offline calibration of 3D laser and camera. Afterwards, Scaramuzza et al. [7] relaxed the condition that the calibration can be conducted by natural images instead of chessboard preparation, whereas the key points still needed to be selected manually. Because the calibration with individual keypoints is easy to generate misalignment, Sergio et al. [9] utilized a circular object to obtain a structure-based calibration. Park et al. [6] adopted a polygon constructed by several keypoints to fulfill a structure alignment. These hand-operated methods can achieve a relatively accurate calibration under controllable condition, whereas need laborious manual operation. Hence, some automatic modules for calibration appeared in recent years.

Automatic calibration: Automatic calibration aims to accomplish an automatic calibration in different scenarios. For instances, Kassir and Peynot [3] automatically extracted the chessboard region in camera and laser sensors, and achieved a keypoint based calibration. Geiger et al. [2] utilized one image and the corresponding laser data to automatically fulfill the calibration with more than one chessboards in the image plane. Scott et al. [8] automatically chosen the natural scenes for a better calibration condition selection and reduced the misalignments. These automatic calibrations are all based on specific circumstances without the consideration of dynamic factors of scenes, and vulnerable to camera jitter. Aiming at this issue, online module begins to attract the attention recently. Within this domain, Levinson and Thrun [5] proposed an online

automatic calibration with an updating for calibration parameters with the latest several frames. However, the calibration parameters are easy to generate bias with the optimal ones because of the improper optimization process.

To this end, inspired by the work of [5], this paper re-formulates the online calibration problem, and using a more reliably spatial structure matching for parameter updating. In the meantime, we rectify the 3D points with the aid of differential inertial measurement unit (IMU), and increase the frequency of 3D laser data as the same as the ones of RGB data. The experimental analysis demonstrates that the proposed method can generate a high-accurate calibration performance even with a tough initialization of calibration parameters. The flowchart of the proposed calibration method is demonstrated in Fig. 2.

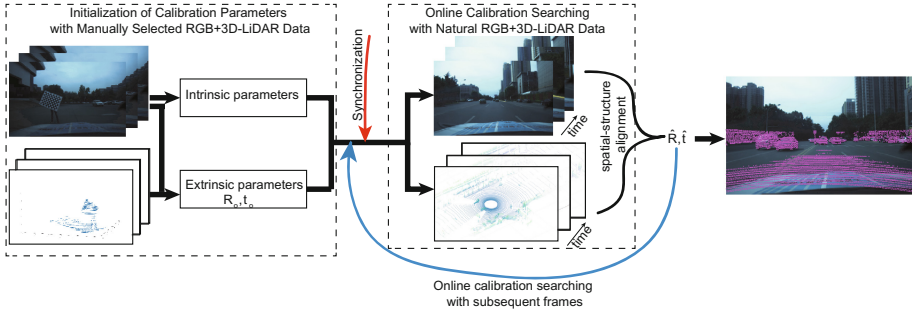


Fig. 2. The flowchart of the proposed calibration method. Firstly, we compute the initialization calibration parameters by manually selecting several images and the corresponding laser data with a chessboard auxiliary. Secondly, the synchronization of RGB and 3D-LiDAR data is conducted for a calibration data preparation. Thirdly, the online high-accurate calibration with a more reliable spatial-structure matching of natural RGB and 3D-LiDAR data is given.

2 Synchronization of RGB and 3D-LiDAR Data

The purpose of the calibration in this work is to automatically determine the six dimensional transformation with a series of image frames and corresponding 3D point sets. Because of the demand of the same object being targeted by two different sensors, this work firstly fulfills the synchronization of the RGB and 3D-LiDAR data, and increases the frame frequency of laser scanning data as the same as the ones of RGB data with an aid of differential internal measurement unit (IMU). IMU in this work provides a 6-dimensional pose state of vehicle, including the location $\mathbf{p} = (x, y, z)$ and its corresponding rotation pose (*roll, yaw, pitch*). Influenced by the scanning module of 3D-LiDAR, all of the 3D points have the different timestamps and pose state. Consequently, the state of each 3D point is distinct. For a calibration task, we need rectify the pose state of the each 3D point as the same as the one of image. Specifically, for a circle of scanning by 3D-LiDAR, we can obtain the starting and ending time of one

round of scanning, and the indexes of the user datagram protocol (UDP) bags with a constant time interval. By these two kinds of information, we can calculate the timestamps and the pose state of each 3D point. Given the timestamps and pose state vector of each image frame, the synchronization is carried out by transforming the pose state of each 3D point into the one of image.

Assume the timestamps of a 3D point and the image is t_0 and t_1 . Denote the location of the 3D point at timestamps of t_0 and t_1 is $\mathbf{p}_0 = (x_0, y_0, z_0)$ and $\mathbf{p}_1 = (x_1, y_1, z_1)$, and the rotation matrix of t_0 and t_1 corresponding the origin pose state is \mathbf{R}_0 and \mathbf{R}_1 , respectively. The transformation of $\mathbf{p}_0 \rightarrow \mathbf{p}_1$ is computed by:

$$\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = \mathbf{R}_1^T \mathbf{R}_0 \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \mathbf{R}_0(\mathbf{p}_1 - \mathbf{p}_0), \quad (1)$$

where \mathbf{R}_0 and \mathbf{R}_1 are calculated by the multiplication of three rotation matrixes respectively in *roll*-axis, *yaw*-axis, and *pitch*-axis. With that, we achieve the synchronization of RGB and 3D-LiDAR Data. One remaining issue is that the frame frequency of RGB data and 3D-LiDAR Data is different. Usually, the capturing frequency of camera is larger. In other words, there is no corresponding 3D points for some video frames. To this end, we take a principle of proximity that we assign the 3D points to the RGB frame with a minimum interval of timestamps of them. As thus, we increase the frequency of 3D LiDAR data as the same as the one of RGB data, which forms the basis for subsequent spatial calibration.

3 Online High-Accurate Calibration

In this section, we will present the online high-accurate calibration. For online calibration, there are three steps: (1) initialization of calibration parameters containing the intrinsic parameter of camera and the extrinsic parameter of camera+3D-LiDAR; (2)online calibration with an adequate spatial alignment optimization. In the following, the initialization of the calibration parameters is firstly described, and then the online high-accurate calibration approach is given.

3.1 Initialization of Calibration Parameters

For the initialization of calibration parameters, it contains intrinsic parameter of camera and extrinsic parameter of camera+3D-LiDAR. As for the intrinsic parameter of camera, this work introduces the commonly work proposed by Zhang [12], where the best intrinsic parameter of camera is selected from 20 gray-scale images with a chessboard auxiliary. With respect to the extrinsic parameter of camera+3D-LiDAR, we extract the image plane of chessboard region in 20 images and the corresponding laser plane. Then, the extrinsic parameters, i.e., translation vector $\mathbf{t} = (\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{z})$ for each point and the corresponding rotate matrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ for *roll*, *yaw*, and *pitch*, are calculated by the Laser-Camera Calibration Toolbox (LCCT) [10].

3.2 Online Calibration

Online calibration aims to update the extrinsic parameter by newly observed camera frames and a series of corresponding laser scans for a better adaptation to scene variation. Inspired by the work of [5], it achieved the online calibration by the edge alignment of several newest camera images and the frames of laser scanning. However, this method is easy to generate bias for the calibration parameters, and the calibration results drift away from the accurate situation. To this end, given a calibration of \mathbf{t} and \mathbf{R} , this work firstly projects the 3D LiDAR sensor onto the image plane, and then formulates online calibration problem as:

$$\max : E_{\mathbf{R}, \mathbf{t}} = \sum_{f=n-w}^n \sum_{p=1}^{|V^f|} V_p^f S_{i,j}^f, \quad (2)$$

where w is the frame number for optimization (set as 9 frames in this work), n is the newest observed video frame, p is the index for 3D point set $\{V_p^f\}_{p=1}^{|V^f|}$ obtained by 3D LiDAR sensor, $S_{i,j}^f$ is the point (x, j) in the f^{th} frame S . The points in both sensors are all the edge points. Similar to [5], the point $S_{i,j}^f$ in image is extracted by edge detection (the method for edge detection is not the focus) appending an inverse distance transformation (IDT), and 3D point $\{V_p^f\}$ is obtained by calculating the distance difference of the scene from the scanner of 3D LiDAR. This formulation is similar to the work of [5], but with a better strategy for optimal calibration searching.

Optimization. For the solving of Eq. 2, this work adopts the grid search algorithm. Specifically, we set the initialization calibration as $\{\mathbf{R}_o, \mathbf{t}_o\}$, which is treated as the optimal calibration $\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\}$ at the beginning of searching. Because there are six values in the transformation of calibration, i.e., x, y, z translations, and *roll*, *yaw*, *pitch* rotations. When performing the grid search with radius 1, there will be $3^6 = 729$ configurations (denoted as perturbations by [5]) of $\{\mathbf{R}, \mathbf{t}\}$ centered around $\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\}$ for the k^{th} grid of searching, denoted as $\{\mathbf{R}_i^k, \mathbf{t}_i^k\}_{i=1}^{729}$. We define $\{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k\}$ as the temporary optimal calibration at the k^{th} grid of searching. With these configurations, this work utilizes them to project the 3D point onto the image plane, and computes $E_{\mathbf{R}_i^k, \mathbf{t}_i^k}$ for each configuration $\{\mathbf{R}_i^k, \mathbf{t}_i^k\}$ at the k^{th} grid of searching. For the temporary optimal calibration $\{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k\}$, we select it by:

$$\{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k\} = \arg \max_{\mathbf{R}_i^k, \mathbf{t}_i^k} E_{\mathbf{R}_i^k, \mathbf{t}_i^k}. \quad (3)$$

To evaluate the quality of $\{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k\}$, the work of [5] defines it as the fraction that the configurations with $E_{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k} > E_{\mathbf{R}_i^k, \mathbf{t}_i^k}$ relative to all the configurations, denoted as $F_{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k}$. Then, they terminate the searching if $F_{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k} - F_{\hat{\mathbf{R}}^{k-1}, \hat{\mathbf{t}}^{k-1}} < \varepsilon$, where ε is a small constant value. Otherwise, $\hat{\mathbf{R}} \leftarrow \hat{\mathbf{R}}^k, \hat{\mathbf{t}} \leftarrow \hat{\mathbf{t}}^k$. However, this quality evaluation of [5] is vulnerable to noise configuration, and easy to drift away from the accurate calibration.

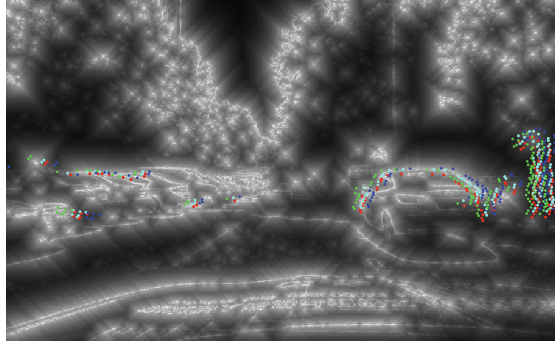


Fig. 3. The demonstration of the spatial-structure alignments under the best calibration and several surrounding configurations. Among them, the points with cyan color, representing the edge points of 3D-LiDAR, have a best spatial alignment with the edge of the image after inverse distance transformation (IDT). In the meanwhile, we also show some configurations neighboring the best one by other colors. From this figure, we can observe that configurations around the best calibration almost generate the similar alignments. Note that this figure should be exhibited with RGB color mode. (Color figure online)

In the end, this paper defines the quality of $\{\hat{\mathbf{R}}^k, \hat{\mathbf{t}}^k\}$ by computing the variance of $\{\mathbf{R}_i^k, \mathbf{t}_i^k\}_{i=1}^{729}$, denoted as $V(E_{\mathbf{R}_i, \mathbf{t}_i^k})$. Note that the variance is calculated after a L2-normalization of $E_{\mathbf{R}_i, \mathbf{t}_i^k}$. We terminate the searching if $V(E_{\mathbf{R}_i, \mathbf{t}_i}) < \tau$, where τ is a small constant for evaluating the smoothness of $E_{\mathbf{R}_i, \mathbf{t}_i^k}$. Otherwise, $\hat{\mathbf{R}} \leftarrow \hat{\mathbf{R}}^k, \hat{\mathbf{t}} \leftarrow \hat{\mathbf{t}}^k$. That means that the smaller value of the $V(E_{\mathbf{R}_i, \mathbf{t}_i})$ has, the better the calibration is. The behind meaning of this strategy is that if we obtain the best calibration $\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\}$, the configurations around $\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\}$ can generate a relative similar value of $E_{\mathbf{R}, \mathbf{t}}$. Taking Fig. 3 as an example, because of the inverse distance transformation, the spatial-structure matchings of RGB and 3D-LiDAR data under the calibrations of $\{\mathbf{R}_i, \mathbf{t}_i\}_{i=1}^{729}$ have a similar results when the optimal calibration is searched. When the optimal calibration is obtained, it is treated as $\{\mathbf{R}_o, \mathbf{t}_o\}$ for online calibration of subsequent image frames and the corresponding series of laser scans. Actually, the calibration in this work accomplishes a truly structure alignment of the data collected by camera and 3D-LiDAR.

In the following, we will give the experiments to validate the performance of the proposed method.

4 Experiments and Discussions

4.1 Dataset Acquisition

The experimental data is collected by an autonomous vehicle named as “Kuafu”, which is developed by the Lab of Visual Cognitive Computing and Intelligent Vehicle of Xian Jiaotong University. In this work, we utilize the equipments

containing a Velodyne HDL-64E S2 LIDAR sensor with 64 beams, and a high-resolution camera system with differential GPS/inertial information. The visual camera is with the resolution of 1920 1200 and a frame rate of 25. In addition, the scanning frequency of the 3DLiDAR is 10 Hz. Thousands of frames for calibration are evaluated.

4.2 Implementation Details

To evaluate the performance of the proposed method, two other calibration approaches are selected. One is the camera-laser calibration in the LCCT [10], and another is the online calibration method of [5]. Because of the dynamic scene and unpredictable camera jitter, it is difficult to obtain the ground-truth of the calibration parameters. In order to give a fair comparison, this work evaluates the performance from two aspects: (1) demonstrating calibration results when searching the optimal calibration by the work of [5] and the proposed method; (2) giving some snapshot comparisons with the calibration parameters by different methods.

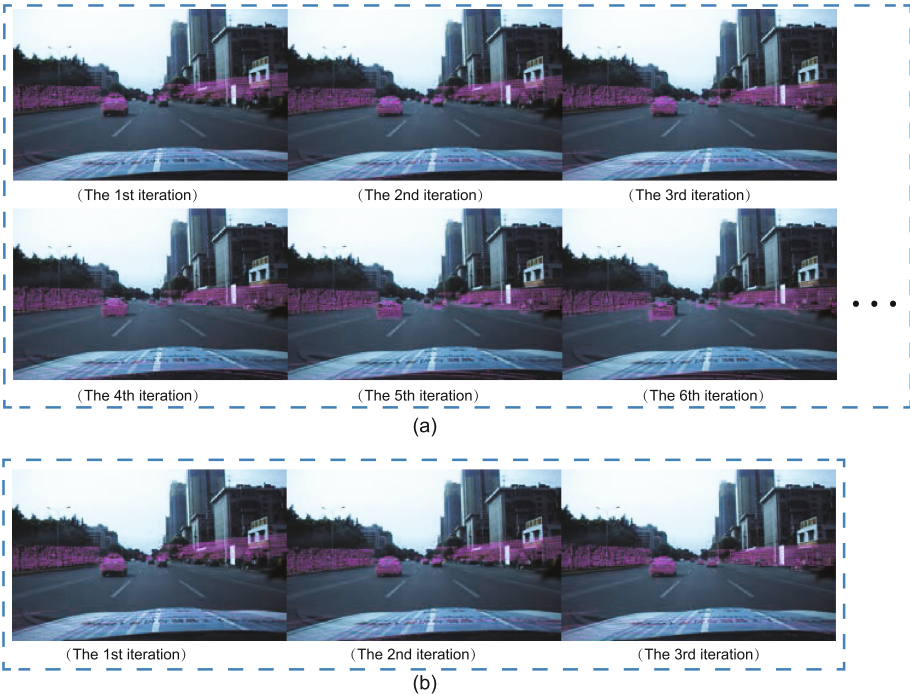


Fig. 4. The iteration steps of (a) the work of [5] and (b) the proposed methods. The suspension points in this figure means that the searching procedure has reached the end.



Fig. 5. The snapshots of typical calibration results. The first column is the results generated by LCCT [10]. The second column is the ones obtained by [5], and the results outputted by the proposed method shown in the third column.

4.3 Performance Evaluation

Figure 4 demonstrates the iteration process for searching the optimal calibration by [5] and the proposed method. From the shown results, we can see that the searching process of [5] drifts away, while our method can obtain a more accurate

calibration. For finding the reason to this phenomena, we have checked each iteration step, and found that the criterion of the terminal condition in [5] is not adequate. That is because they terminate the searching when the fraction that all other configurations generate lower value for Eq. 2 than the center configuration remains unchanged. This terminal condition is vulnerable to noise calibration with larger value for Eq. 2. Actually, this phenomena is universal within the work [5]. As for the proposed method, we emphasis that all the configurations in the grid of searching should generate large values for Eq. 2 when the best calibration is searched, which is a truly spatial-structure alignment for RGB+3D-LiDAR data. Therefore, the best calibration is outputted after iterating.

In addition, we also give several snapshots for the calibration comparison, shown in Fig. 5. From this figure, it is manifestly that the proposed method obtains the best calibration for the demonstrated images. Actually, these observation is general in the comparing methods.

From the above analysis, the superiority of the proposed method is validated.

4.4 Discussions

When performing the calibration, there is an universal phenomena that the accuracy of the calibration is different for the entire scene. It is common in all the calibration methods. That is to say when the near scene is correctly calibrated, the calibration results for the far scene may appear skewing, and vice versa. The main reasons are that: (1) the RGB and Laser data themselves have distortion to some extent; (2) the calibration parameters are searched from the entire scene, which cannot avoid the influence of the regional distortion data in RGB+3D-LiDAR sensors. This problem may be tackled from a local calibration view in the future.

5 Conclusion

This paper presented an online high-accurate calibration method for RGB+3D-LiDAR sensors in the autonomous driving circumstance. Through a synchronization, we rectified the 3D points gathered by 3D-LiDAR data with a aid of differential inertial measurement unit (IMU), and increased frame frequency of laser data as the same as the one of visual camera. Then, this work obtained an online high-accurately calibration via a more reliable spatial-structure matching of RGB and 3D-LiDAR data. The superiority of the proposed method was verified in the experiments.

Acknowledgement. This work is supported by the National Key R&D Program Project under Grant 2016YFB1001004, and also supported by the Natural Science Foundation of China under Grant 61603057, China Postdoctoral Science Foundation under Grant 2017M613152, and is also partially supported by Collaborative Research with MSRA.

References

1. Buehler, M., Iagnemma, K., Singh, S.: The DARPA Urban Challenge: Autonomous Vehicles in City Traffic. Springer, Heidelberg (2009). <https://doi.org/10.1007/978-3-642-03991-1>
2. Geiger, A., Moosmann, F., Car, Ö., Schuster, B.: Automatic camera and range sensor calibration using a single shot. In: Proceedings of the IEEE Conference on Robotics and Automation, pp. 3936–3943 (2012)
3. Kassir, A., Peynot, T.: Reliable automatic camera-laser calibration. In: Proceedings of the 2010 Australasian Conference on Robotics and Automation (2010)
4. Le, Q.V., Ng, A.Y.: Joint calibration of multiple sensors. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3651–3658 (2009)
5. Levinson, J., Thrun, S.: Automatic online calibration of cameras and lasers. In: Proceedings of the Robotics: Science and Systems (2013)
6. Park, Y., Yun, S., Won, C.S., Cho, K., Um, K., Sim, S.: Calibration between color camera and 3D LiDAR instruments with a polygonal planar board. *Sensors* **14**(3), 5333 (2014)
7. Scaramuzza, D., Harati, A., Siegwart, R.: Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4164–4169 (2007)
8. Scott, T., Morye, A.A., Pinis, P., Paz, L.M.: Choosing a time and place for calibration of LiDAR-camera systems. In: Proceedings of the IEEE Conference on Robotics and Automation, pp. 4349–4356 (2016)
9. Sergio, A.R.F., Fremont, V., Bonnifait, P.: Extrinsic calibration between a multi-layer LiDAR and a camera. In: Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 214–219 (2008)
10. Unnikrishnan, R., Hebert, M.: Fast extrinsic calibration of a laser rangefinder to a camera. Carnegie Mellon University (2005)
11. Zhang, Q., Pless, R.: Extrinsic calibration of a camera and laser range finder (improves camera calibration). In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2301–2306 (2005)
12. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
13. Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., Hu, S.: Traffic-sign detection and classification in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2110–2118 (2016)
14. Zhu, Z., Lu, J., Martin, R.R., Hu, S.: An optimization approach for localization refinement of candidate traffic signs. *IEEE Trans. Intell. Transp. Syst.* **18**(11), 3006–3016 (2017)